

A Framework for an HSI Downselection Tool

Phillip M. Cunio

M.L. Cummings

Massachusetts Institute of Technology*

Prepared for Dr. T. Skrmetti, Rite Solutions, Inc.

HAL2009-03

July 2009



<http://halab.mit.edu>

e-mail: halab@mit.edu

*MIT Department of Aeronautics and Astronautics, Cambridge, MA 02139

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE JUL 2009		2. REPORT TYPE		3. DATES COVERED 00-00-2009 to 00-00-2009	
4. TITLE AND SUBTITLE A Framework for an HSI Downselection Tool				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, Humans and Automation Laboratory, Cambridge, MA, 02139				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 26	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Contents

Acronym List	3
Introduction.....	4
Motivation and Context	4
Theoretical Development.....	6
The Overall SITHE Process.....	6
Ratings in SITHE	8
Phase A: Preparatory Analytical and Rating Phase	9
Phase B: Guided Human Decision-Making.....	9
Binary Questions.....	9
Lifecycle Questions	11
Trade Space Questions.....	14
Tool Performance Questions.....	16
Phase C: Automatic Calculation	17
Phase D: Human Visualization	17
Case Studies	19
Conclusions.....	21
Future Work	21
Preparatory Analytical and Rating Phase (Phase A):	21
Guided Human Decision Phase (Phase B):.....	22
Automatic Calculation Phase (Phase C):	22
Human Visualization Phase (Phase D):	22
Acknowledgments.....	23

References	24
Appendix A: Full Expanded Hierarchy Figure	25

ACRONYM LIST

DSS: Decision Support System
 GFE: Government Furnished Equipment
 HSI: Human-Systems Integration
 SITHE: Systems Integration Tool for HSI Evaluation
 SRR: System Requirements Review
 PDR: Preliminary Design Review
 CDR: Critical Design Review

INTRODUCTION

This technical report describes the concept and development of SITHE, the Systems Integration Tool for HSI Evaluation. SITHE is a framework for selecting tools to be used in evaluating complex technical systems in terms of Human-Systems Integration, or HSI.

HSI, or Human-Systems Integration, is the process of integrating people, technology, and an organization at a systems level, with full consideration given to the human requirements of the user (Booher, 2003). HSI focuses on the human aspects of system definition, development, and deployment, and integrates considerations related to personnel, training, human factors, habitability, and other human-related concerns into the overall systems acquisition process (US Department of Defense, 2004). HSI is a field of interest to researchers in academia and industry because, although systems continue to grow more complex, they have not achieved the level of autonomy that would permit them to operate successfully without humans either in or on the loop. Humans are still an essential component of most complex systems, especially when the context of operation for the complex system is subject to uncertainty, as in military applications. However, HSI as a broad field can encompass a large number of types of interaction between humans and systems, including but not necessarily limited to supervisory control, mechanics and ergonomics of control operation, and visualization and decision support.

The universe of tools for HSI (including hardware, software, processes, and techniques used to evaluate HSI aspects of complex systems) is already large and growing quickly. Many HSI tools are developed for research purposes only, or in an ad-hoc fashion for specific projects, and as such there is no such thing as a standard catalogue of HSI tools. In addition, the need to consider downstream competencies such as flexibility, robustness, and usability, is increasing as HSI systems become more complex. Thus the HSI cost-benefit trade space is ever increasing, making it difficult for decision makers to determine if and to what degree a system actually meets some pre-specified HSI criteria.

MOTIVATION AND CONTEXT

HSI has become an area of increasing concern and interest to the general engineering community since roughly the end of the Second World War, and as a result, there exists a plethora of tools for HSI evaluation. The current size of the HSI “toolbox” numbers at least approximately 275 tools (at least that many have been collected in a database by Rite Solutions, Inc., in a related project). Many more tools may exist, especially when considering that “tool” can be defined as any software, paper, or other product or process that allows evaluation of an HSI-related characteristic of a system.

Some of these tools are catalogued and well-documented, especially those sold commercially, but there are many more tools used and developed for research purposes, often solely by the researchers themselves, and the general community that might otherwise be interested in these tools often has no or very limited knowledge of them. Given this situation, it is difficult for an engineer, manager, or other decision-maker to know how to evaluate the HSI-related aspects of a system using tools from the toolbox as it currently exists.

As an example, consider the case of a senior decision-maker for a complex technical project. If the decision-maker (perhaps an acquisitions manager or a systems engineering lead) wants to assess the HSI-related aspects of a system, in light of evaluating and possibly acquiring the system, he or she requires tools to analyze the system's HSI-related aspects. The first problem to solve is which of the tools are most appropriate to use in evaluating the system?

The decision support framework described in this paper, SITHE (the Systems Integration Tool for HSI Evaluation), is intended as a downselection aid, meant to help solve the problem of selecting tools to evaluate a complex system in terms of HSI. Ideally, after making this downselection, the decision-maker is left with a smaller and more efficient set of tools to use. The decision-maker can then acquire and apply these tools, and then evaluate the HSI-related state of a complex technical system, and possibly decide between two or more different complex systems. This approach is not meant to identify the best-of-breed or most widely known tools for HSI, but rather those most appropriate for application in a given context by a given decision-maker.

This report primarily describes a framework for carrying out downselection to a final set of HSI tools useful for evaluation of a complex system, with the specific processes and methods identified here being only examples of the general processes and methods. While this paper does describe a specific process for downselection to a final set of HSI tools, it should be seen as an initial solution, i.e., a starting point rather than a final answer.

The approach is also meant to be generalizable. The selection of tools for any activity from a large set can be addressed by an application of SITHE. Application of a very large HSI tool set may not be practical either due to manning or monetary considerations, so how to go from a large to a small set in a principled, objective, and repeatable manner is the generally purpose of SITHE. To this end, SITHE takes a database of information on the set of all existing tools and applies to it the downselection process to arrive at a tool set which is comprehensible and useful to a decision-maker. Although this means that SITHE tends toward the abstract (i.e., it is a decision-support system intended for use on other decision-support systems), it is nonetheless a significant aid in the first step of the process of evaluating complex technical systems.

THEORETICAL DEVELOPMENT

Downselection is the process of comparing members of a large set and eliminating some of them in order to arrive at a smaller set. The process of downselection is often carried out cyclically in order to shrink very large sets to manageable size. In order to downselect from a large set, it is necessary to evaluate and compare members of that set. Several schemes for applying rules to the set exist, generally rooted in classical decision theory. For the SITHE downselection process, an elimination-by-aspects process coupled with multi-criteria weightings are combined to generate a final candidate tool set, from which human users select the final tool set. Elimination by aspects is the process of removing members from a set based on undesirable aspects of members (Lehto, 1997). Every member with an undesirable aspect is eliminated given a list of undesirable aspects that is sequentially applied to the set in order to shrink it.

The elimination-by-aspects process used for SITHE relies on a list of binary questions to make an initial screening of the total available tool set. Tools which do not pass a binary criterion are eliminated from the final tool set. The list of binary questions deals with different aspects of the tools, and is applied sequentially to the total available tool set to shrink it. The remaining tools are then rated according to several other multi-criteria aspects, and the results are displayed to the decision-maker. These results are not binary, but fall along a continuum and are split among several axes. The result is a trade space, where some tools score higher along some axes than others. At the final stage, the human decision-maker applies judgment to make a selection of tools for the final tool set.

The following description of the overall SITHE process includes details on methods for all steps in the procedure.

THE OVERALL SITHE PROCESS

The SITHE process is a mix of human and automation effort, with two human agencies. The upstream user (or agency), who gathers and collates information for the database that drives later phases of the process, interacts with the data asynchronously. The downstream user, who either is or acts for the decision-maker in charge of a complex technical system that is being evaluated by the SITHE framework, is the ‘customer’ who must answer questions about the specific downselection objectives. The automation acts on the information created by the upstream and downstream users, according to pre-programmed rules, and generates a smaller set of possible tools for use from the database, acting as a filtering screen for the human downstream user. The downstream user reviews this list of tools, can conduct sensitivity analyses if desired, and makes the final decision about which tools (and how many tools) to use for the final evaluation of the HSI system.

The SITHE process includes four major phases. These four phases are Phase A: preparatory analytical effort (by the upstream user), Phase B: guided downselection effort (by the downstream user), Phase C: automatic tool calculations, and Phase D: final human decisions, aided by some visualization (carried out by the downstream user). The overall process appears in Figure 1, and these four phases are delineated within the overall process in Figure 2. Figure 3 shows the roles played by the upstream and downstream users in the overall process.

An example of using the SITHE process would include a senior decision-maker as the downstream user and a small staff of HSI experts as the upstream users. The upstream users would create a database and conduct an initial evaluation of the tools in the database. This database would be processed by the downstream user, who would answer several series of questions, which would generate a binary filter screen and tool rating weights. The rating weights would be automatically aggregated with the tool evaluation results and the binary filter screen, which would generate a reduced list of tools and scores associated with aspects of these tools. The final phase allows the senior decision-maker to examine the reduced tool list and select the tool or tools desired for the final tool set.

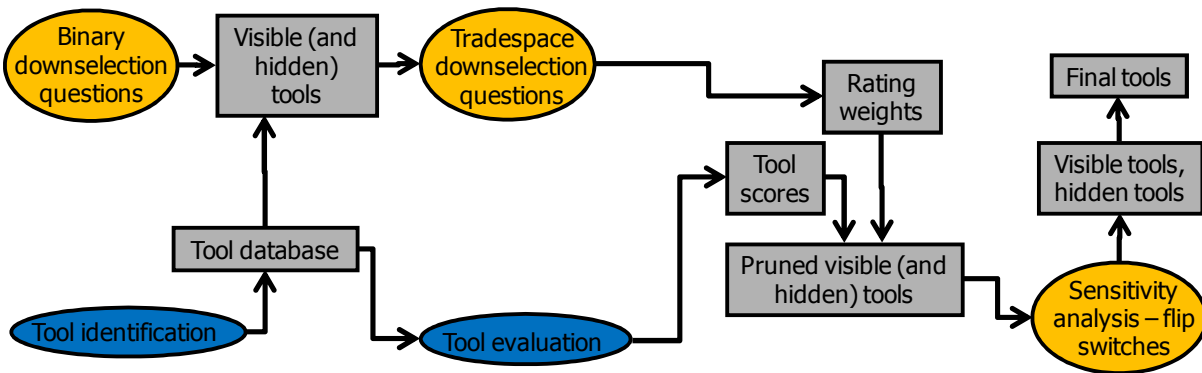


Figure 1. The SITHE process.

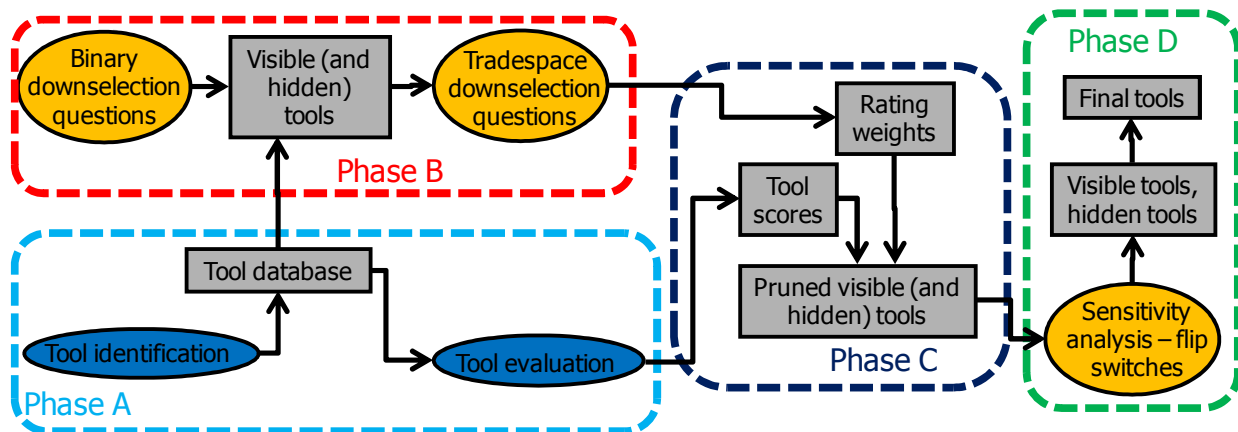


Figure 2. Phases of the SITHE Process.

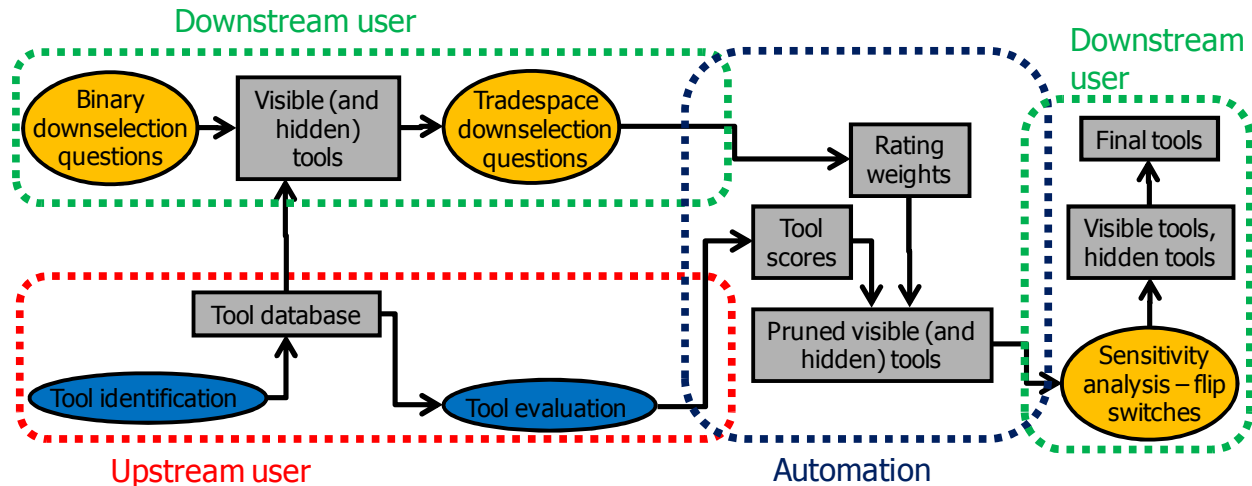


Figure 3. Upstream and Downstream Users in the SITHE Process.

RATINGS IN SITHE

In order to generate data which can be used for elimination by aspects, it is necessary to create by some means ratings or rankings for all the aspects which will be used during the downselection process. As initially developed for this report, SITHE relies on two sets of ratings: one generated by the upstream user, who has a higher degree of subject matter expertise and a more broad perspective, and one generated by the downstream user, who has less subject matter expertise, but also has a specific complex technical project in mind for evaluation. This general procedure of amalgamating information from a more objective viewpoint and a more subjective viewpoint is meant to increase the robustness of the results of using SITHE.

Although the specific method of amalgamation of the data from upstream and downstream users may be changed in future iterations of SITHE, this initial version uses three types of ratings for tools, which are intended to address three major questions of interest. The first question addresses whether the tool can be used at all. This general question is answered by means of a binary question list, so that only tools that make it through the binary filter system are retained. The second question addresses whether the tools can be used for the phase of the project in question. The downstream user may wish to explore a range of tools for various systems engineering lifecycle phases (e.g, the set of metrics for the conceptual design phase is quite different from the set of metrics for final test and evaluation). The third question addresses the degree to which each tool will benefit the downstream user. Downstream users have particular areas of interest, and each tool again has particular areas to which it is applicable. The higher the overlap between the two, the more likely it is that the tool will be retained and placed into the final tool set that the downstream user could use to evaluate the HSI system.

A distinction should be made between the binary questions and trade space questions. Binary questions serve to eliminate tools with certain aspects (or lack of certain aspects) entirely. Trade

space questions acknowledge that all tools have the aspects in question to various degrees, and quantify those degrees. Binary questions are meant as a rapid-filtering device, as they have no gray area between the two extremes - tools either meet a criterion or they don't. The binary questions serve to rapidly shrink the set of all tools to a manageable size. Trade space questions are meant to quantify as well as qualify, and are able to show the downstream user more detailed information and allow for sensitivity analysis.

For this initial iteration of SITHE, the trade space ratings from both the upstream user and the downstream user are created on a scale from 1 to 5, with 5 indicating the highest degree of performance on a given aspect (for the upstream user) or the highest importance (for the downstream user). The upstream user, who is assumed to possess a high degree of subject matter expertise, examines each tool and rates how well it performs for each aspect. The downstream user rates how important performance in each aspect is specifically for his or her purposes. The downstream user is not expected to have the time or expertise to evaluate each tool separately, and so the downstream user's total rating workload is much lower.

The following sections of this report describe sequentially the four phases of the overall SITHE process.

PHASE A: PREPARATORY ANALYTICAL AND RATING PHASE

The first step of the SITHE process is carried out by the upstream user. In this phase, a database of information on HSI tools is compiled. The information collected in this phase will later be operated upon in the downselection process. Although much information can be collected from existing databases, such as the online Directory of Design Support Methods (US Department of Defense, 2009), it must also be organized into aspects to facilitate downselection. For the initial iteration of SITHE, an Excel[®]-based database was used.

In this phase, the upstream user also generates ratings for each tool in terms of its aspects for the binary question list (i.e., by answering the question, does this tool have this aspect, yes or no?), and generates an assessment of the phases of the systems engineering lifecycle and the technical areas of interest to which each tool is applicable, as well as the performance of the tool according to several key metrics.

PHASE B: GUIDED HUMAN DECISION-MAKING

BINARY QUESTIONS

The first part of the guided downselection phase of the SITHE process includes a set of binary questions. The easiest way to screen a large set is to assign every element in the set a binary score and then eliminate every member with the non-desirable score. Examples of this include threshold values, where every element with a score below a certain level is eliminated, and yes-

no dichotomies, where every element is retained or eliminated according to whether or not it falls in a particular category. At the most basic level, binary evaluations answer the question “Does this tool have characteristic X?” Trade space evaluations answer the question “To what degree does this tool have characteristic X?” Binary questions are essentially a trade space question with the entire answer space split into just two regimes.

In this phase, all the questions are answered on a binary scheme, evaluated as a Yes (1) or No (0). A representative list in Figure 4 is presented to the downstream user, and matches a similar list in the database, previously populated by the upstream user. The binary list for the downstream user addresses the needs and wants of the downstream user in evaluating the system. The binary list for the upstream user addresses tool characteristics. For example, a tool either requires subject matter expertise to use or does not, and either requires standard or special hardware or does not, etc. The result is two separate types of filtering tables, as seen in Table 1 and Table 2, which are generated and applied by the automation.

I am willing to hire a Subject Matter Expert in order to use a tool
I am willing to upgrade or buy new hardware in order to use a tool
I am willing to upgrade or buy new software in order to use a tool
I want easily available technical support from the vendor
I want easily available technical support from a third party
I want to be able to purchase a tool off-the-shelf
I want to avoid regulatory compliance efforts associated with a tool (e.g., I want an ITAR-free tool)
I want to use a tool that has met certain certification standards
I am willing to pay to get a tool officially certified
I need to use a particular tool vendor
I need to use a tool that will be Government-Furnished Equipment (GFE)
I need to use a tool for which my sponsors/customers will pay
I need to pay less than a specific amount for a tool

Figure 4. Initial Representative Binary Question List.

Table 1. Filtering Table for Tool Needs.

	Upstream user: Tool needs X	
Downstream user: I am willing to provide/buy X	<u>Yes</u>	<u>No</u>
<u>Yes</u>	<i>Keep tool visible</i>	<i>Keep tool visible</i>
<u>No</u>	<i>Hide tool</i>	<i>Keep tool visible</i>

Table 2. Filtering Table for Tool Wants.

Downstream user: I want a tool with X	Upstream user: Tool has X	
	<u>Yes</u>	<u>No</u>
<u>Yes</u>	<i>Keep tool visible</i>	<i>Hide tool</i>
<u>No</u>	<i>Keep tool visible</i>	<i>Keep tool visible</i>

As seen in Table 1 and Table 2, the answers generated by the upstream and downstream users fall into four possible combinations. Either both answer with a 1, both answer with a 0, or there is a mix. The two tables correspond to two different types of situations, although notably, both tables include only one of four possible cases where the tool in question is hidden, which means the tool is removed from the downstream user’s list of possible tools.

For binary questions related to additional requirements of the tool (e.g., additional hardware), Table 1 is applicable. This table eliminates any tool which requires a cost that the downstream user is not willing to pay. For binary questions related to characteristics of the tool itself (for instance, available technical support from a third party), Table 2 is applicable. Any tool which does not offer a benefit that a downstream user wants is eliminated.

Note that every tool must pass through every filtering table sequentially. Therefore, a longer list of binary questions is, in general, a stronger filtering tool. The results of automated filtering processes are visible to the downstream user, as they govern which tools are visible and which are hidden. Any tool which passes through the entire sequence of binary questions without triggering a “Hide” response remains visible.

Some binary questions could potentially be expanded into trade space questions which indicate degree of match instead of a simple yes or no, especially those that relate to non-beneficiary stakeholders (e.g., regulatory issues, GFE, and certification). Some downstream users will be able to (and desire to) answer some binary questions with more fine detail than others. An example of this is the binary question about regulatory issues: a question ranking the degree of difficulty associated with regulatory issues for a particular tool on a scale from 1 to 5 could be an additional trade space question. Ultimately, the upstream user populates the field of possible binary questions, but the downstream user should have the ability to move questions into the trade space section, if he or she decides the level of granularity is too coarse.

LIFECYCLE QUESTIONS

While the first set of binary questions address tool attributes (the what), the second set of binary questions address the when, through considering the phases of the systems engineering lifecycle in which a tool is applicable. Lifecycle phase applicability questions inquire whether the

downstream user is interested in examining the system during various portions of the systems engineering lifecycle. For the purposes of SITHE, the systems engineering lifecycle is split into five parts (three major sections and two interstitial sections), and the downstream user simply inputs a Yes or No response for each of these five sections. The canonical waterfall model of the systems engineering lifecycle, adapted as in Figure 5, is used as a template. The three major divisions of the lifecycle (concept, prototype, production) are indicated in Figure 6.

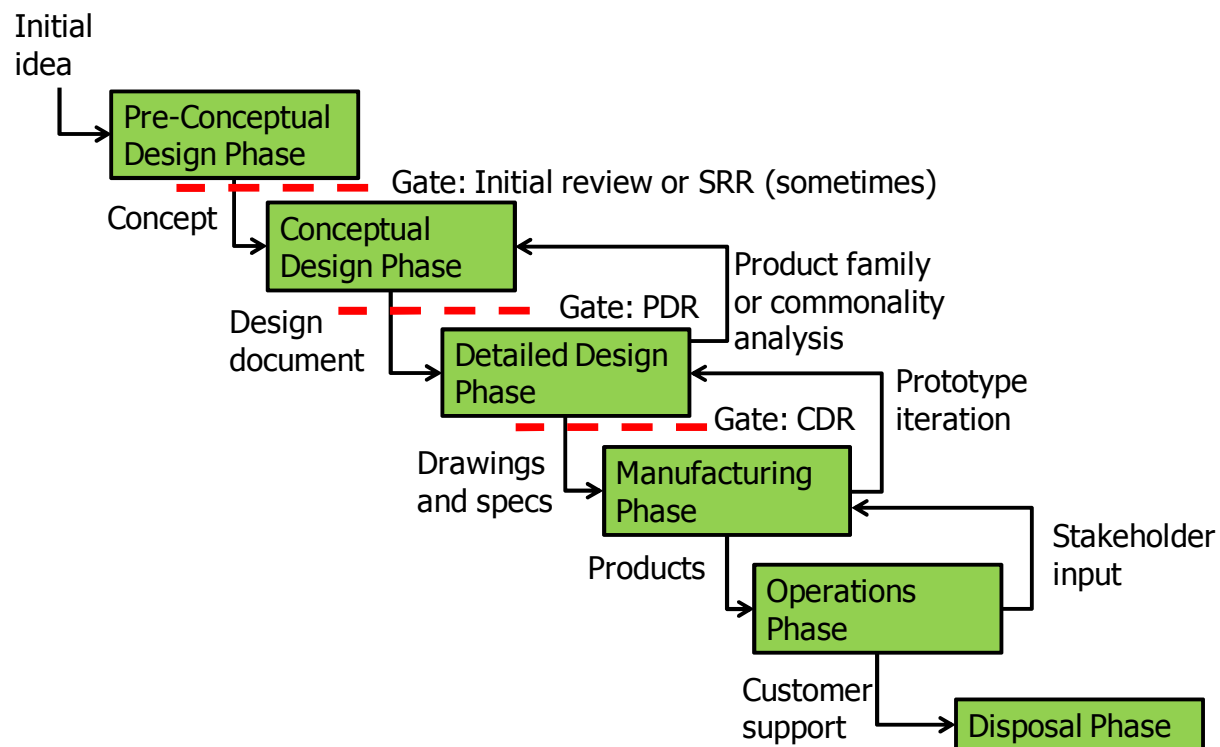


Figure 5. Systems Engineering Lifecycle Model (A. P. Sage and W. B. Rouse, 2003).

The purpose of including lifecycle phases is allowing the downstream user to use SITHE as an exploratory tool, with which he or she might look into later stages of projects, or at the full lifecycle of speculative projects.

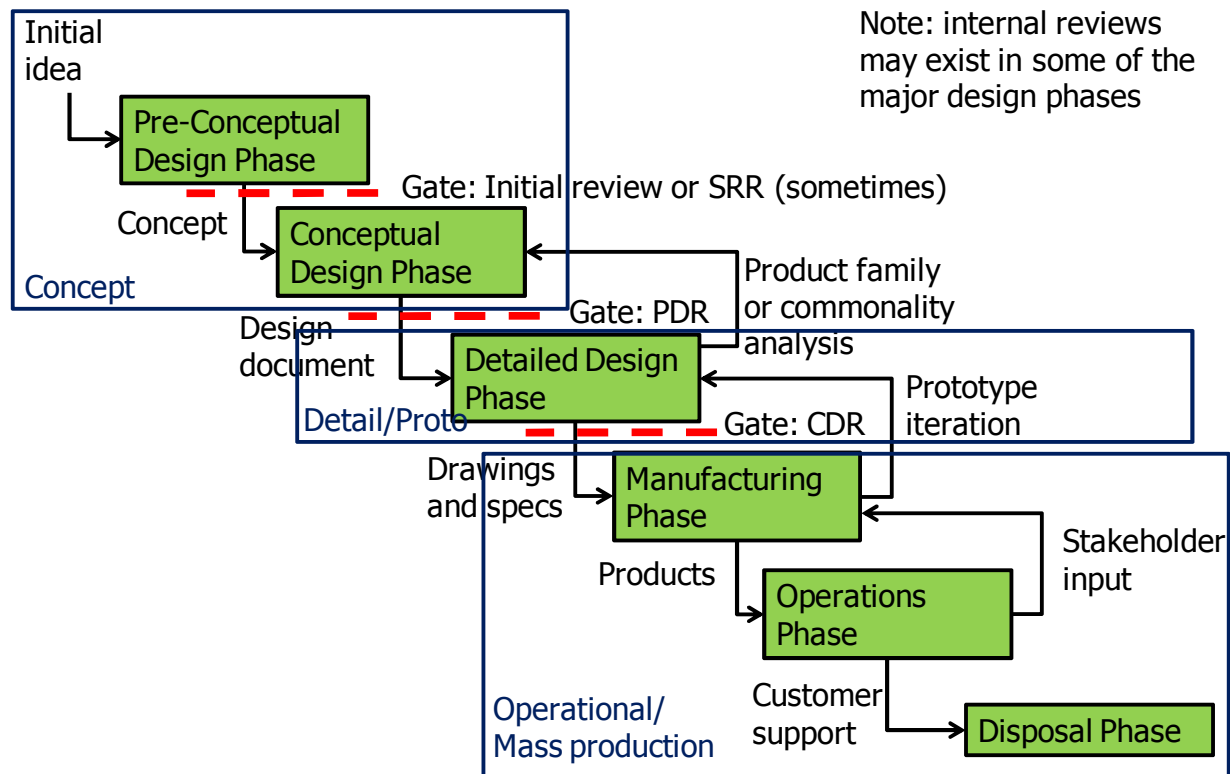


Figure 6. Systems Engineering Lifecycle Model Showing SITHE Life Cycle Segments.

The downstream user selects lifecycle phases of interest via the interface seen in Figure 7. The downstream user inputs zero to indicate no interest in a lifecycle phase, and one to indicate interest. The lifecycle phases are identified as Early, Middle, and Late, with two interstitial phases between the three main phases, and each phase is also described by the typical products for a design in that stage of the lifecycle, as well as the next upcoming major development gates typically expected. These descriptors help the downstream user identify lifecycle phases of interest.

Phase	Early lifecycle	Early-Middle	Middle lifecycle	Middle-Late	Late lifecycle
	0	0	1	1	1
Typical Products	Slides, drawings		Prototype, components		Mass-produced units
Next Gate	Concept Review, Critical Review		Production Review, Prototype Demonstration		Operations Review, Redesign Review

Figure 7. Downstream User's Interface for Lifecycle Interest Indication.

Note that the downstream user's interface for indicating interest in lifecycle phases is a part of the main interface for the downstream user, which appears in full in Figure 13.

TRADE SPACE QUESTIONS

Trade space questions are the final segment of Phase B. In order to develop a cost-benefit trade space for HSI tools, a taxonomy of three dimensions was developed, related to the familiar concept of the iron triangle in which cost, schedule, and performance are interrelated. Trades favoring one dimensional axis must weaken the others (Seaver, 2009). The iron triangle can also be expanded into an iron tetrahedron by adding a risk axis (which is usually included on the performance axis).

The three major dimensions specific to HSI are performance, cost, and usability. The three major dimensions of performance, cost, and usability are not called axes (as the term misleadingly implies that they are orthogonal), but rather taxa. The taxa and subtaxa are also interconnected in subtle ways, unlike axes. Figure 8 shows these three taxa as the initial levels of a hierarchy based on the iron triangle concept. Note that the schedule dimension is not considered here, as the point of SITHE is to evaluate complex systems at various points in the general lifecycle, as opposed to set schedules.

The downstream competency dimension (also known as the “ility” dimension) is set apart in Figure 8 because it describes characteristics of the system which affect its performance in ways that are primarily evident during later phases of its lifecycle. Given that SITHE is intended to apply primarily to the field of HSI, usability is a key concern. However, while not considered here, this hierarchy could easily be extended to include other downstream competencies, such as manufacturability, flexibility, or robustness.

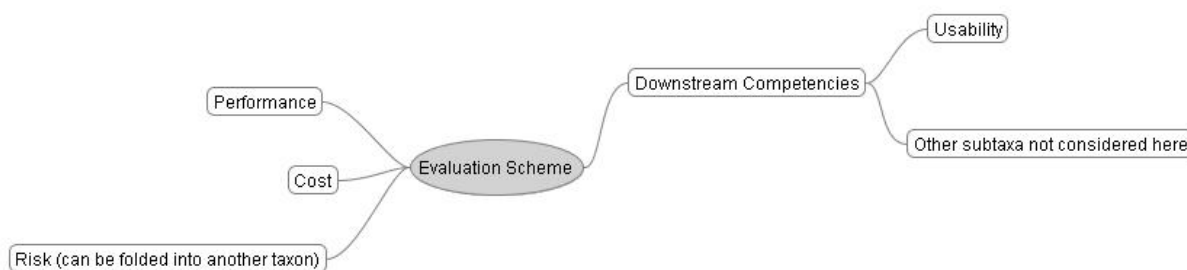


Figure 8. High-level View of Taxa Hierarchy.

The hierarchy in Figure 8 can be expanded, with some exemplars seen in Figure 9, and a representative full hierarchy is in Appendix A. Many of the existing HSI tools (as taken from the database compiled by Rite Solutions, Inc.) are mapped into the hierarchy, according to the subtaxon to which they best correspond. This proof-of-concept mapping clearly indicates that some branches of the hierarchy are glutted with tools (such as cost), while others such as subtaxa in usability are apparently completely empty of tools. In and of itself, this taxa hierarchy mapping process can indicate where gaps in the set of all existing tools lie.

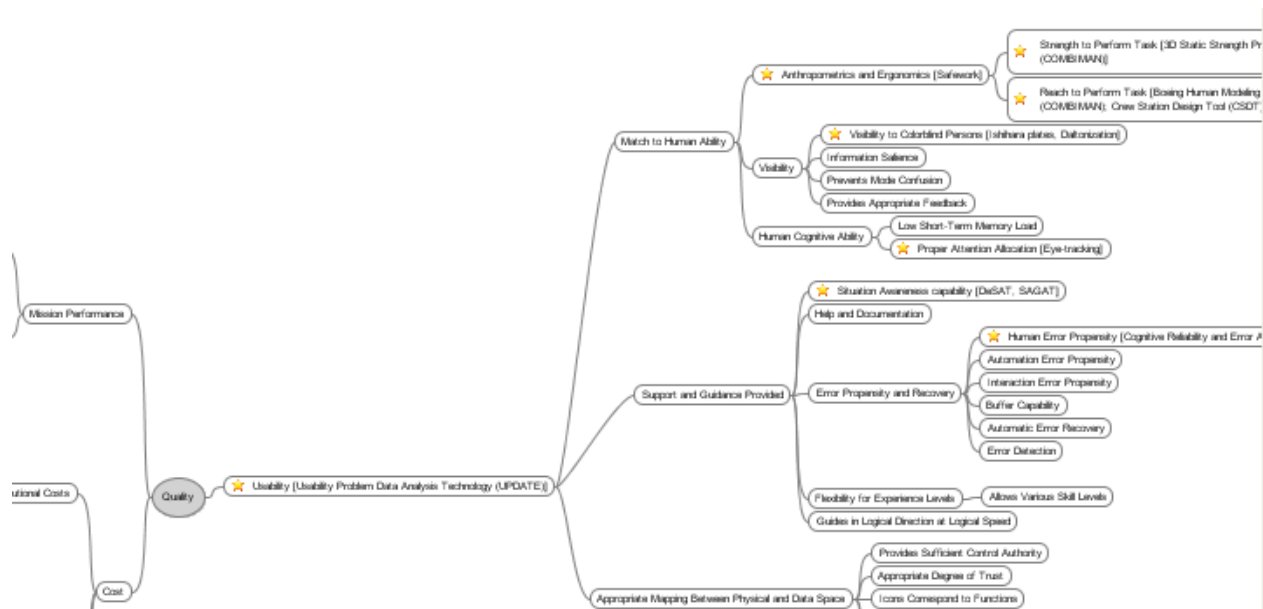


Figure 9. Expanded Partial View of Hierarchy.

The taxa hierarchy is applied in the SITHE process to determine which specific technical areas are of importance to the downstream user. The downstream user steps through the branches of the hierarchy, assigning a rating on a scale of 1 to 5 of the importance of each taxon and subtaxon. Figure 10 shows a series of ratings assigned by a downstream user to some of the taxa and subtaxa in the hierarchy.

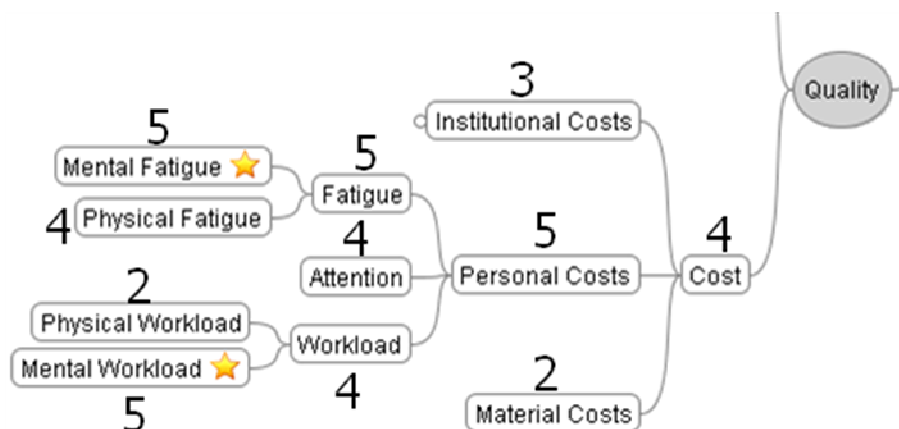


Figure 10. Ratings Assigned by Downstream User to Specific Technical Areas.

Figure 11 shows how these ratings are entered into SITHE. The hierarchy is converted into an Excel table, and the ratings associated with each taxon and subtaxon are recorded in the table. Figure 11 shows the ratings from Figure 10 being entered into the SITHE table. Note that only a small section of the hierarchy is actually seen, and some sections of the hierarchy that were not

visible in Figure 10 (which can be actively expanded and contracted) do appear in Figure 11. The downstream user is in the process of entering ratings into the unexpanded subtaxa in Figure 11.

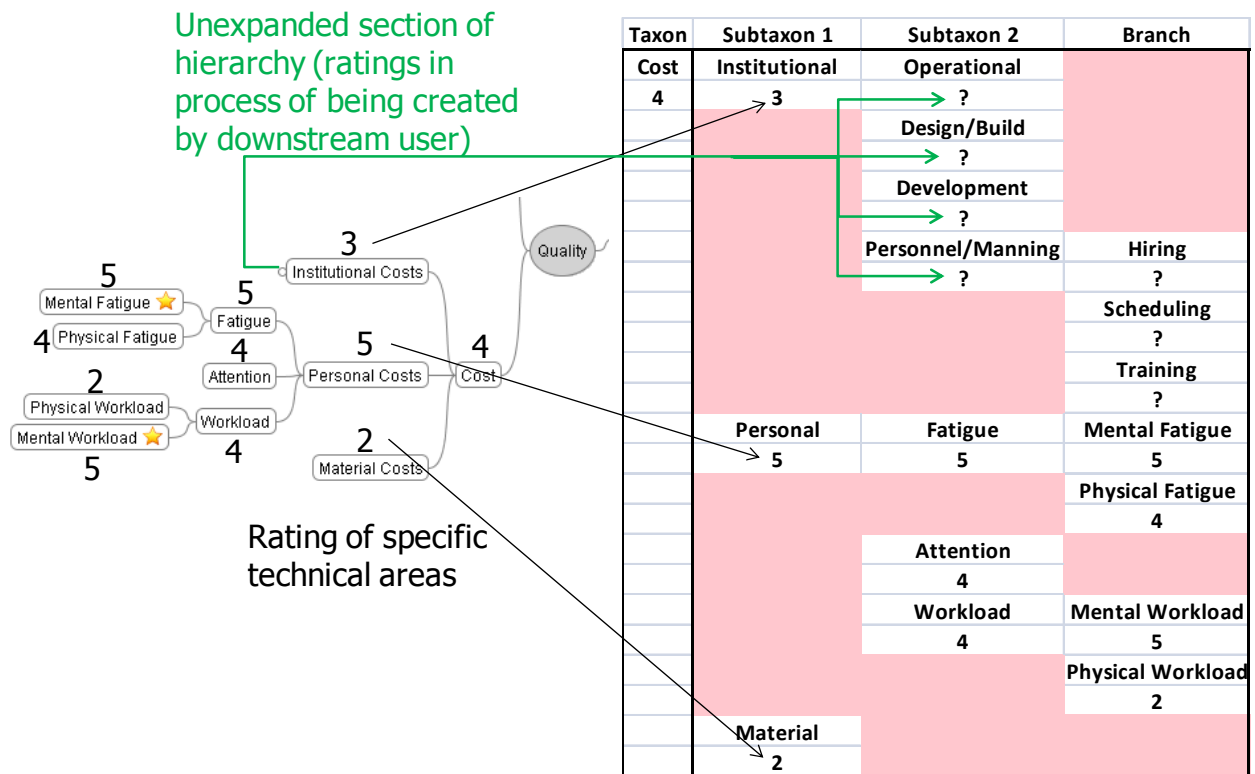


Figure 11. Mapping of Specific Technical Area Ratings into SITHE.

The initial version of SITHE uses this rating as a filtering factor for the lifecycle applicability, one of the scores which appear in Figure 12, simply by matching the rating given to each specific technical area by the downstream user to the tools which are mapped into that area. For example, in Figure 11, the downstream user has assigned the Mental Workload subtaxon a rating of 5, meaning that any tool mapped into that taxon by the upstream user (such as the NASA Task Load Index) will have a rating of 5. However, the Physical Workload subtaxon has been assigned a rating of 2 by the downstream user, meaning that it is unimportant to the downstream user, and any tool mapped into it will be filtered out from further consideration. Future iterations of the hierarchy may include cross-links between the scores given to specific subtaxa and the taxa which contain them, as well as allow for the fact that some tools can be mapped into multiple subtaxa. More complex rating algorithms are possible, although the initial version of SITHE uses a simple one.

TOOL PERFORMANCE QUESTIONS

The final set of trade space questions answered by the downstream user relates to the performance of the tools themselves. Five dimensions for evaluating HSI tools are adapted from previous work by (Donmez, 2008). These five dimensions are Construct Validity, Measurement Efficiency, Comprehensive Understanding, Statistical Efficiency, and Experimental Constraints.

Construct Validity describes how well the tool describes what is actually occurring. Measurement Efficiency describes the ease with which the tool can be used to carry out measurements. Comprehensive Understanding describes how well the tool's output affords a complete picture of what is happening. Statistical Efficiency describes how well the tool's output lends itself to statistical analysis. Experimental Constraints describe the external factors related to the use of this tool.

These metrics represent a means of evaluating the tools on a few characteristics that may be of importance to decision-makers, but will also vary among tools. The tools are rated from 1 to 5 by both the upstream and the downstream users, and these scores are amalgamated into a final score. As with the technical area trade space questions, the upstream user makes an objective assessment of the tool's abilities, while the downstream user makes a subjective assessment of the importance to his or her project of the tool's abilities.

PHASE C: AUTOMATIC CALCULATION

The automatic calculation phase takes the output of the previous phases, as well as information from the upstream and downstream users, and amalgamates it. Binary questions are used to make tools visible (if they pass all binary filters) or hidden (if they do not pass one or more binary filters). Only the visible tools remain available for the downstream user to select.

The visible and hidden tools all also receive scores according to the degree to which their applicable lifecycle phases overlap with those of interest to the downstream user (weighted by the importance of the specific HSI-related technical area that the tool addresses, as assigned by the downstream user), and a set of scores created by the tools' scores on the five performance metrics, weighted by the downstream user's ratings of the importance of those five performance metrics. These scores are all made visible to the downstream user during Phase D of the SITHE process, described next.

PHASE D: HUMAN VISUALIZATION

The current method of visualizing the tool scores uses an Excel[®] plot, as shown in Figure 12. This interface is a preliminary development interface, and development of a more intuitive interface is left for future work.

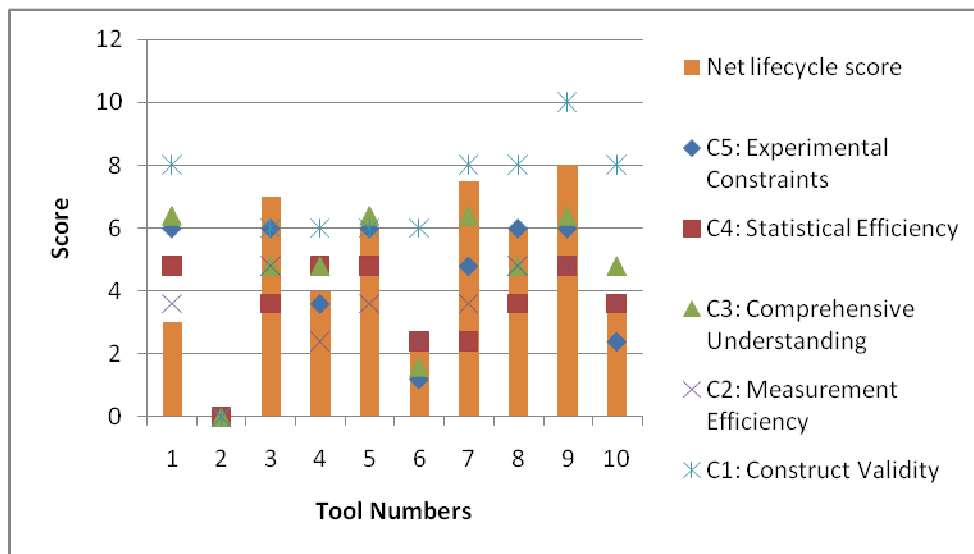


Figure 12. Example of Preliminary SITHE Visualization.

In the Excel[®] visualization scheme, every tool in the database (or some subset of the tools in the database) is listed along the horizontal axis. Several types of scores (normalized so that maximum scores in each scoring axis are equal) are plotted on the vertical axis. The tools which have a score of zero are just dots along the horizontal axis. These scores are set to zero because the tools to which they correspond do not pass through the binary filter. A simple adjustment of the binary filter causes the tools to “jump up” from the horizontal axis and appear on the vertical axis, to be compared to the rest of the toolset as appropriate. The non-visibility or visibility of a tool’s scores on the vertical axis corresponds to whether the tool has been made hidden or not. This representation allows decision makers the ability to conduct sensitivity analyses and “what-if” comparisons, so that they can adjust their personal weighting criteria to determine what set of tools would be available if various parameters were changes.

The solid bars show the extent to which a tool can be used for this project. The height of the bars represents the degree to which the phases of the systems engineering lifecycle in which the tool is applicable (as rated by the upstream user) correspond to the phases of the lifecycle in which the downstream user is interested. The bars are also filtered by the downstream user’s rating of the specific technical area to which the tool is applicable, as seen in Figure 11. A bar reaching the top of the plot in Figure 12 (again, the scores are normalized) would indicate a tool that is applicable to every phase of the lifecycle, and addresses a specific technical area rated high in importance by the downstream user. The height of a bar is directly proportional to its lifecycle applicability, filtered by its specific technical area applicability. This method is meant to convey a means for comparing the applicability of one tool to another at a glance. Bars with a height of zero are not weighted highly enough by their specific technical area of applicability to be considered important (a threshold value of 2 on the 1 to 5 scale was used as the cutoff for being

considered important). Otherwise, the height of the bar scales directly with the tool's score. For example, in Figure 12 Tool 6 is not very applicable, while Tool 7 and Tool 9 are very applicable.

The final set of icons on the plot in Figure 12 shows the individual scores for the tool performance metrics for each tool. One type of icon corresponds to each of the five metrics for tool performance (construct validity, measurement efficiency, comprehensive understanding, statistical efficiency, and experimental constraints). Each tool along the horizontal axis will have an individual (although not necessarily unique) set of these scores.

The visualization of these separate parameters allow humans, such as the downstream user, to flip binary switches, adjust the phases of the systems engineering lifecycle that are considered of interest, and see how each individual tool might benefit him or her along several characteristics. This information allows the downstream user to make a final decision as to which tools should be included in the final toolset (note that the tools which appear high-scoring in SITHE are not necessarily the only possible choices, just those predicted to be among the best choices). The faculty of human decision-making is thus combined with the power of automated scoring and analysis to select the most optimal set of tools for the purposes of the senior decision-maker who acts as the downstream user.

CASE STUDIES

As initial validation for SITHE, three case studies were conducted. Users were given a brief explanation of the concept and workings of SITHE, and were then guided by a human through the scoring processes. Finally, the toolset initially recommended by SITHE was examined by the users, who then selected a final preferred toolset. Then, the results were compared to the users' pre-existing expectations of what evaluation tools were appropriate for their systems.

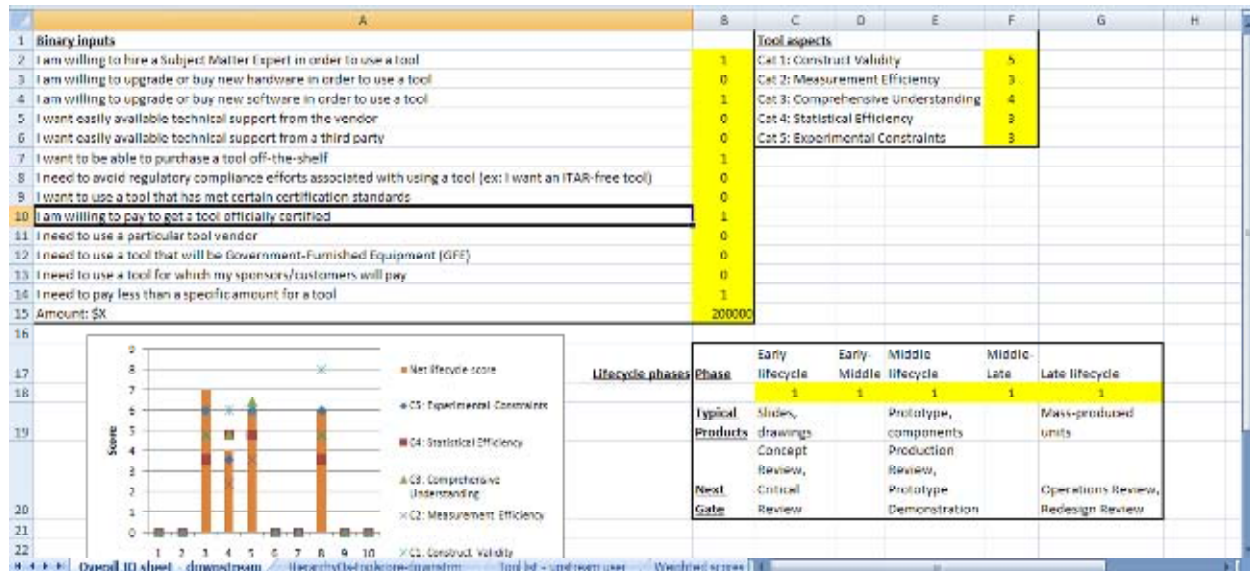


Figure 13. SITHE User Interface for Pilot Tests.

The user interface for these usability evaluations included a graphic of the hierarchy (as seen in Appendix A), and an overall input/output screen (Figure 13). Pilot users were given a representative sample of ten tools and asked to play the role of the downstream user. Pilot users could thus change their entries to the various scoring areas (highlighted in Figure 13) while actively watching the output plot for changes in the toolset, facilitating their application of human judgment during Phase D of the pilot tests.

Before the pilot tests, one of the authors served in the role of upstream user, generating the necessary ratings for the tools in the representative database. The same author also guided the pilot users through the process.

Two HAL students were drafted as pilot users, and were asked to use SITHE to evaluate complex technical systems with which they had extensive familiarity. Both students had prior experience in using and developing decision support systems (Carrigan, 2009) (Massie, 2009). In both cases, the users were able to use SITHE effectively, and the final recommended toolsets matched their expectations as to system evaluation. The toolsets initially marked as potentially desirable by SITHE were either larger than or equal to the final selected toolsets (after the human input and decision-making in Phase D of the process), indicating that SITHE is probably effective, at least initially, in predicting tools that will be of interest.

The third case study was conducted with industry experts from Rite Solutions, Inc. The SITHE tool was presented in the same way, and the industry experts were asked to consider the design and evaluation of a system with which they were very familiar. While the initial result was that no tool from the representative database included in the test version of SITHE was deemed

feasible, this was actually remarked upon as realistic. The fractured state of HSI is such that finding tools that can be practically used in industry is very difficult, and often compromises with initial answers to binary questions must be made in order to allow any tools to be feasible.

During the pilot tests, the industry experts, and to a lesser extent the HAL users, wanted to see the highest level of the automated scoring process in SITHE. That is, they looked into the parts of the initial SITHE tool where the binary filters were applied, and noted details – for example, which binary filters were actively hiding a particular tool, or how exactly the upstream user had rated the applicability of a particular tool to the phases of the systems engineering lifecycle. Upon seeing some of these details, the industry experts were able to return to the main input/output screen and make decisions about where changes to their answers could be made. Eventually, they were able to select a tool that they found useful and viable. SITHE showed a score for this tool that was less than the highest, but still registered on the scoring scale.

The test with industry experts indicated that some further upgrades, such as icons for seeing the factors which rule out any one tool, may be useful, and that an expanded database would also help. However, the SITHE tool itself proved useful overall, and passed its initial usability evaluation in that a final decision on a tool set was able to be made within a reasonable amount of time.

CONCLUSIONS

Initial usability tests indicate SITHE is a useful and valid tool for the process of downselecting to a small set of HSI tools. Given a large set of tools that do not always apply for the different lifecycle phases and maturity of desired systems, SITHE is able to bring users to a final toolset with which they feel comfortable within a reasonable amount of time. In addition, SITHE's initial recommendations and filters were seen as reasonable by pilot testers. SITHE is ready for further development or more detailed trials with additional users and additional systems to require evaluation.

FUTURE WORK

The SITHE process encompasses four major phases, and each of these phases shows potential for additional research questions. A brief summary of some of these potential additional research questions appears below, grouped by the relevant phase (refer to Figure 2).

PREPARATORY ANALYTICAL AND RATING PHASE (PHASE A):

- How does the identity/experience of the person conducting the preparatory effort influence the final tool choices?

The biases of the upstream and downstream users could impact the ratings given to tools, which affect the remainder of the process. Further work is needed to determine if there is a means of compensating for possible biases, for example, using people from two different organizations and obtaining concurrence between them.

GUIDED HUMAN DECISION PHASE (PHASE B):

- Do the order and type of trade space/binary downselection questions presented affect the final ratings?

The order in which questions from the binary, lifecycle phase applicability, specific technical area, and tool performance metrics question sets may affect the ratings placed on each by the downstream user, due to acclimatization or learning effects. These effects may be present with respect to the order of individual questions and the order of the overall question sets.

- Do the order and type of trade space/binary questions affect the final tool selection?

In a similar fashion, the ratings generated by users may, via their own bias, affect the final tool selection that downstream users make. The case may also be that the ratings are variable, but the final tool selections are not.

AUTOMATIC CALCULATION PHASE (PHASE C):

- Should downstream users be allowed to adjust upstream users' ratings of tools?
- In a related hypothesis, do users apply SITHE in such a way to justify an existing bias towards or against a tool?

Given that downstream users may introduce individual biases, either consciously or not, to obtain a preferred toolset, it may be appropriate to regulate the extent to which downstream users have power over the ratings generated by upstream users. On the other hand, the understanding of an experienced downstream user may lend itself to rewriting the upstream user's ratings in order to drive the process in such a way that it truly facilitates a decision by the downstream user.

HUMAN VISUALIZATION PHASE (PHASE D):

- Does the visualization method for final human downselection affect the final tool choices? Does it affect the perception of utility, either for the tools in the final tool set or for SITHE itself?

The ways in which visualization, including the hidden/visible effect created by the binary filtering tables as well as the multidimensional scoring plot shown in Figure 12, affects the final outcome of the process is an important area of investigation. The proper visualization tools to facilitate decision-making and to allow easy use of SITHE must be developed and implemented.

ACKNOWLEDGMENTS

This project was sponsored by the Office of Naval Research and Rite Solutions, Inc.

REFERENCES

Sage, A.P. & W. B. Rouse, ed. (2003). *Handbook of Systems Engineering and Management*. New York: Wiley-Interscience.

Donmez, B., Pina, P., & M.L. Cummings. (2008). Evaluation criteria for human-automation performance metrics. *Performance Metrics for Intelligent Systems Workshop*. Gaithersburg, MD.

Booher, H. R. (2003). *Handbook of Human Systems Integration*. Wiley-Interscience.

Carrigan, G. P. (2009). The Design of an Intelligent Decision Support Tool for Submarine Commanders. *SM Thesis*. Cambridge, MA: MIT Engineering Systems Division.

Lehto, M. R. (1997). Ch. 37: Decision Making. In E. Gavriel Salvendy, *Handbook of Human Factors and Ergonomics, 2nd Edition*. New York: John Wiley and Sons, Inc.

Cummings, M.L. & P. Pina (2008). A Metric Taxonomy for Human Supervisory Control of Unmanned Vehicles. *Association for Unmanned Vehicle Systems International*. San Diego.

Massie, A. E. (2009). Designing a Graphical Decision Support Tool to Improve System Acquisition Decisions. *SM Thesis*. Cambridge, MA: MIT Department of Aeronautics and Astronautics.

Pina, P. & M. L. Cummings (2008). Identifying Generalizable Metric Classes to Evaluate Human-Robot Teams. *Metrics for Human-Robot Interaction Workshop, 3rd Annual Conference on Human-Robot Interaction*. Amsterdam.

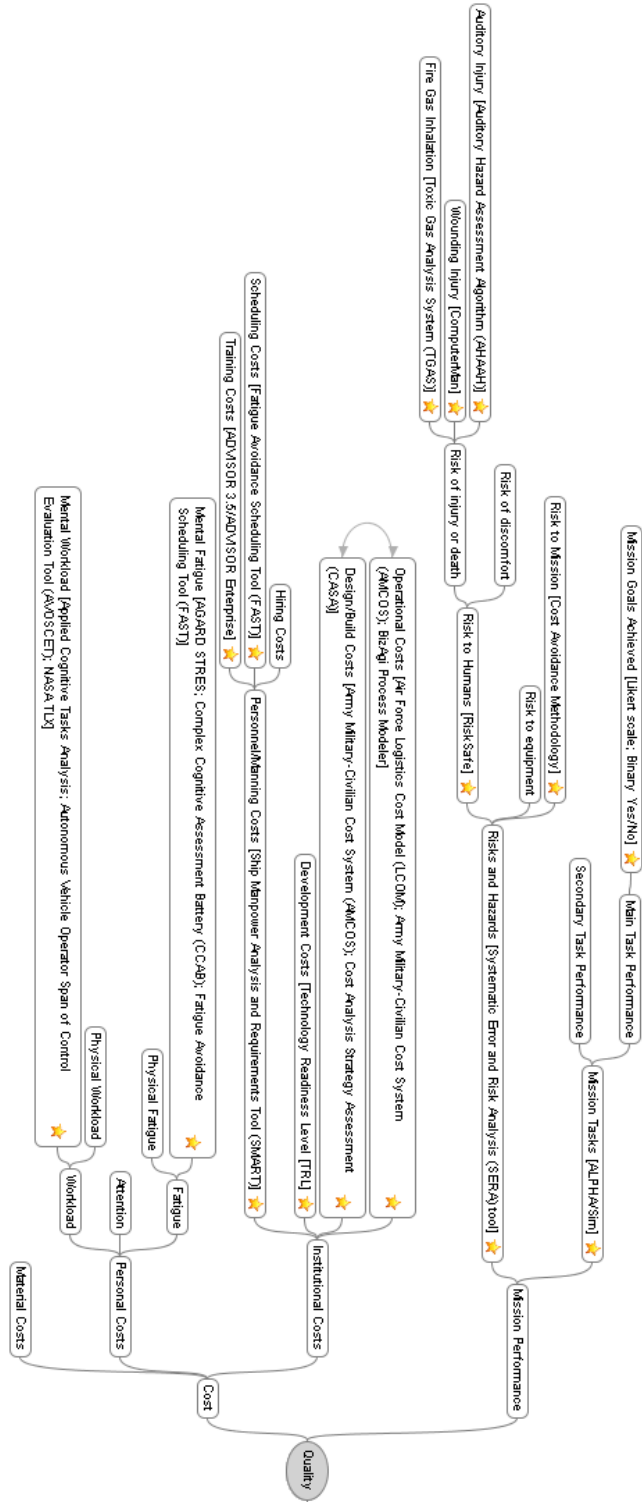
Seaver, D. *Respect the Iron Triangle*. Retrieved July 1, 2009, from IT Today:
<http://www.ittoday.info/Articles/IronTriangle.htm>

US Department of Defense. (2004). *Chapter 6: Human Systems Integration (HSI)*. Retrieved June 30, 2009, from AT&L Knowledge Sharing System:
https://akss.dau.mil/dag/Guidebook/IG_c6.1.asp

US Department of Defense. (2009). *Directory of Design Support Methods*. Retrieved June 30, 2009, from Defense Technical Information Center, San Diego:
<http://www.dtic.mil/dticasd/ddsm/>

APPENDIX A: FULL EXPANDED HIERARCHY FIGURE

Left side:



Right side:

